INFINITE SIDON SEQUENCES

JAVIER CILLERUELO

ABSTRACT. We present a new method to obtain infinite Sidon sequences, based on the discrete logarithm. We construct an infinite Sidon sequence $\mathcal{B}$, with $\mathcal{B}(x) = x^{\sqrt{2}-1+o(1)}$. Ruzsa proved the existence of a Sidon sequence with similar counting function but his proof was not constructive.

Our method generalizes to B_h sequences: For all $h \geq 3$, there is a B_h sequence $\mathcal{B}$ such that $\mathcal{B}(x) = x^{\sqrt{(h-1)^2+1}-(h-1)+o(1)}$.

1. INTRODUCTION

In 1932, Simon Sidon asked P. Erdős about the growing of those sequences $\mathcal{B}$ with the property that all sums $b+b',\ b \leq b',\ b,b' \in \mathcal{B}$ are distinct. Erdős named them Sidon sequences. Sidon had found one with $\mathcal{B}(x) \gg x^{1/4}$ and Erdős observed that the greedy algorithm, described below, provides another with $\mathcal{B}(x) \gg x^{1/3}$.

Starting with $b_1 = 1$, let us define b_n to be the smallest integer greater than b_{n-1} and such that the set $\{b_1, \dots, b_n\}$ is a Sidon set. Since there are at most $(n-1)^3$ distinct elements of the form $b_i+b_j-b_k,\ 1 \leq i,j,k \leq n-1$, it is then clear that $b_n \leq (n-1)^3+1$ and the counting function of the sequence $\mathcal{B}$ generated by this algorithm (the greedy algorithm) certainly satisfies $\mathcal{B}(x) \geq x^{1/3}$.

Erdős conjectured that for any $\epsilon > 0$ should exist a Sidon sequence with $\mathcal{B}(x) \gg x^{1/2-\epsilon}$, but the sequence given by the greedy algorithm was, for almost 50 years, the densest example known of an infinite Sidon sequence. That was until that, in 1981, Atjai, Komlos and Szemeredi [1], in a seminal paper where they invented a key lemma about graphs, proved the existence of an infinite Sidon sequence such that $\mathcal{B}(x) \gg (x \log x)^{1/3}$. They wrote:

The task of constructing a denser sequence has so far resisted all efforts, both constructive and random methods. Here we use a random construction for giving a sequence which is denser than the above trivial one.

It was a surprise when, in 1998, Ruzsa [5] overcome the barrier of the exponent 1/3, proving the existence of an infinite Sidon sequence $\mathcal{B}$ with $\mathcal{B}(x) = x^{\sqrt{2}-1+o(1)}$. The

2000 *Mathematics Subject Classification.* 11B83.

Key words and phrases. Sidon sets, B_h sequences, Probabilistic method, Discrete logarithm.

This work was supported by Grant MTM 2005-04730 of MYCIT (Spain).

starting point of Ruzsa's approach was the sequence $\{\log p\}$ where p runs over all the prime numbers. Probably, as Gowers [3] pointed out, Ruzsa had first observed that the set

$$\mathcal{A} = \left\{ \lfloor n\log(4p/\sqrt{n})\rfloor\,,\ \sqrt{n}/4 < p \le \sqrt{n}/2,\ p \text{ prime}\right\},$$

or some variant, is a Sidon set in $[1,n]$ with size $|\mathcal{A}| \gg \sqrt{n}/\log n$.

Ruzsa's proof is not constructive. For each $\alpha \in [1,2]$ he considered a sequence $\mathcal{B}_\alpha = \{b_p\}$ where each b_p is built using the binary digits of $\alpha \log p$. What Ruzsa proved is that for almost all $\alpha \in [1,2]$ the sequence $\mathcal{B}_\alpha$ is nearly a Sidon sequence in the sense that removing not too many elements from the sequence it is possible to destroy all the repeated sums that eventually appear.

Here we present a new method to construct infinite Sidon sequences. It is inspired by the finite Sidon set

$$\mathcal{A} = \{\log_g p :\ p \text{ prime}\ ,\ p \le \sqrt{q}\},$$

where g a generator of $\mathbb{F}_q^*$ and $\log_g x$ denotes the discrete logarithm of x in $\mathbb{F}_q^*$. The set $\mathcal{A}$ is indeed a Sidon set in $\mathbb{Z}_{q-1}$ with size $|\mathcal{A}| = \pi(\sqrt{q}) \sim 2\sqrt{q}/\log q$. Despite the simplicity of this construction we have not seen it previously in the literature.

To warm up we construct first an infinite Sidon sequence $\mathcal{B} = \{b_p\}_{p\in\mathcal{P}}$ indexed with all the prime numbers with an easy explicit expression for the elements b_p. This is the first time that an infinite Sidon sequence $\mathcal{B}$ with $\mathcal{B}(x) \gg x^\delta$ for some $\delta > 1/3$ is constructed explicitly. Theorem 1.1, where we state this result, is weaker than Theorem 1.2, but we have included it as a separated theorem because the simplicity of the construction.

Theorem 1.1. *The sequence $\mathcal{B} := \mathcal{B}_{\overline{q},c} = \{b_p\}_{p\in\mathcal{P}}$ constructed in section 2 is, for $c = \frac{3-\sqrt{5}}{2}$, an infinite Sidon sequence with*

$$\mathcal{B}(x) = x^{\frac{3-\sqrt{5}}{2}+o(1)}.$$

With a more delicate argument we can construct a denser infinite Sidon sequence $\mathcal{B} = \{b_p\}_{p\in\mathcal{P}*}$. Now the set of indexes is not the whole set of the prime numbers, as in the previous construction, but the set $\mathcal{P}^*$, the survived primes after we remove a thin subset of the primes that we can describe explicitly.

Theorem 1.2. *The sequence $\mathcal{B} = \{b_p\}_{p\in\mathcal{P}^*}$ constructed in subsection 2.1 is an infinite Sidon sequence with*

$$\mathcal{B}(x) = x^{\sqrt{2}-1+o(1)}.$$

Note that the exponent of the counting function in the explicit construction of Theorem 1.2 is the same that Ruzsa obtained in his random construction. The algorithm introduced to construct the Sidon sequence in Theorem 1.2 is efficient in the sense that only $O(x^{\sqrt{2}-1+o(1)})$ elementary operations are needed to list all the elements $b_p \le x$.

Our approach also generalizes to B_h sequences, that is, sequences such that all the sums $b_1 + \cdots + b_h,\ b_1 \le \cdots \le b_h$ are distinct. To deal with theses cases, however, we need to introduce a probabilistic argument so that the proof of the following theorem is not constructive.

Theorem 1.3. *For any $h \ge 3$ there exists an infinite B_h sequence $\mathcal{B}$ with*

$$\mathcal{B}(x) = x^{\sqrt{(h-1)^2+1}-(h-1)+o(1)}.$$

The exponents in Theorem 1.3 are greater than $1/(2h-1)$, those given by the greedy algorithm for B_h sequences. It should be mentioned that R. Tesoro and the author [2] have proved recently Theorem 1.3 in the cases $h = 3$ and $h = 4$ using a variant of Ruzsa's approach which makes use of the sequence $\{\theta(\mathbf{p})\}$ of arguments of the Gaussian primes $\mathbf{p} = |\mathbf{p}|e^{2\pi i\theta(\mathbf{p})}$ instead of the sequence $\{\log p\}$ considered by Ruzsa. However, that proof does not extend well to all h.

In the last section we present an alternative method to construct infinite Sidon sequences. It has the same flavor than the construction described in section 2 but the irreducible polynomials in $\mathbb{F}_2[X]$ play the role of the prime numbers in the set of positive integers. The finite version of this idea is the following. If $q := q(X)$ is an irreducible polynomial in $\mathbb{F}_2[X]$ with $\deg q = T$, the set

$$\mathcal{A} = \{x(p) :\ X^{x(p)} \equiv p \pmod{q},\ p \text{ irreducible},\ \deg(p) < T/2\}$$

is a Sidon set in $[1, 2^T]$ with size $|\mathcal{A}| \gg 2^{T/2}/T$.

We present an sketch of how to reprove Theorems 1.1, 1.2 and 1.3 using this alternative approach.

2. The construction

Let $\overline{q} = \{q_j\}$ be a given infinite sequence of prime numbers satisfying $2^{2j-1} < q_j \le 2^{2j+1}$ for all $j \ge 1$. We choose, for each j, a primitive root g_j of $\mathbb{F}_{q_j}^*$. Fix $c,\ 0 < c < 1/2$, and consider a partition of the prime numbers $\mathcal{P} = \cup_{k\ge 2}\mathcal{P}_k$, where

$$\mathcal{P}_k = \left\{ p \text{ prime} :\ \frac{2^{c(k-1)^2}}{k-1} < p \le \frac{2^{ck^2}}{k} \right\}.$$

Let us define the sequence $\mathcal{B}_{\overline{q},c} = \{b_p\}_{p\in\mathcal{P}}$ as follows: for $p \in \mathcal{P}_k$, set

$$b_p = \sum_{1\le j\le k} x_j(p)(4q_1)\cdots(4q_{j-1}), \tag{2.1}$$

where $x_j(p)$ is the solution of the congruence

$$g_j^{x_j(p)} \equiv p \pmod{q_j}, \qquad q_j + 1 \le x_j(p) \le 2q_j - 1$$

and define $x_j(p) = 0$ for $j > k$.

Next we state some properties of the sequence $\mathcal{B}_{\overline{q},c}$.

Proposition 1. *If* $b_{p_1}+b_{p_2}=b_{p_1'}+b_{p_2'}, \quad b_{p_1}>b_{p_1'}\geq b_{p_2'}>b_{p_2}$ *then*

i) *there exist* $k_2,k_1,\ k_2\leq k_1$ *such that* $p_1,p_1'\in\mathcal{P}_{k_1},\ p_2,p_2'\in\mathcal{P}_{k_2}$.
ii) $p_1p_2\equiv p_1'p_2' \pmod{q_1\cdots q_{k_2}}$
iii) $\quad p_1\equiv p_1' \pmod{q_{k_2+1}\cdots q_{k_1}}$ *if* $k_2<k_1$.
iv) $(1-c)k_1^2<k_2^2<\frac{c}{1-c}k_1^2$.

Proof. Let us consider the following well known fact: *Given an infinite sequence of positive integers* $r_1,\cdots,r_j,\cdots$ *(the base), any non negative integer can be written, in an only way, in the form*

$$y_1+y_2r_1+y_3r_1r_2+\cdots+y_jr_1\cdots r_{j-1}+\cdots$$

with digits $0\leq y_j<r_j,\ j\geq 1$.

Then we observe that the digits of b_p in the base $4q_1,\dots,4q_j,\dots$ are

$$b_p=x_kx_{k-1}\cdots x_2x_1.$$

Since $x_j(p_1)+x_j(p_2)<4q_j$, we have that

$$x_j(p_1)+x_j(p_2)=x_j(p_1')+x_j(p_2')$$

for all j.

By construction, $p_1\in\mathcal{P}_{k_1}$ and $p_2\in\mathcal{P}_{k_2}$ where k_1 is the largest j such that $x_j(p_1)+x_j(p_2)\geq q_j+1$ and k_2 is the largest j such that $x_j(p_1)+x_j(p_2)\geq 2q_j+2$. This observation proves part i).

For $j\leq k_2$ we have that

$$g_j^{x_j(p_1)+x_j(p_2)}\equiv g_j^{x_j(p_1')+x_j(p_2')} \pmod{q_j}.$$

Thus $p_1p_2\equiv p_1'p_2' \pmod{q_j}$ for all $j\leq k_2$ and then $p_1p_2\equiv p_1'p_2' \pmod{q_1\cdots q_{k_2}}$.

If $k_2<k_1$, for $k_2+1\leq j\leq k_1$ we have that

$$g_j^{x_j(p_1)}\equiv g_j^{x_j(p_1')} \pmod{q_j}.$$

Thus $p_1\equiv p_1' \pmod{q_j}$ for all $k_2+1\leq j\leq k_1$ and then $p_1\equiv p_1' \pmod{q_{k_2+1}\cdots q_{k_1}}$.

Part ii) and the inequalities on p_i and q_j yield

$$\frac{2^{c(k_1^2+k_2^2)}}{k_1k_2}\geq|p_1p_2-p_1'p_2'|\geq q_1\cdots q_{k_2}>2^{1+3+\cdots+(2k_2-1)}=2^{k_2^2}\implies k_2^2<\frac{c}{1-c}k_1^2.$$

In particular it implies that $k_2<k_1$ and we can apply part iii), which gives

$$\frac{2^{ck_1^2}}{k_1}\geq|p_1-p_1'|\geq q_{k_2+1}\cdots q_{k_1}>2^{(2k_2+1)+\cdots+(2k_1-1)}=2^{k_1^2-k_2^2}\implies k_2^2>(1-c)k_1^2.$$

□

Proposition 2. *The counting function of the sequence $\mathcal{B}_{c,\overline{q}} = \{b_p\}$ defined above satisfies $\mathcal{B}_{c,\overline{q}}(x) = x^{c+o(1)}$.*

Proof. The first observation is that distinct primes p, p' provide different $b_p, b_{p'}$. Indeed, if $b_p = b_{p'}$ then $p, p' \in \mathcal{P}_k$ for some k and $p \equiv p' \pmod{q_1 \cdots q_k}$. Thus, $2^{ck^2} \geq |p - p'| \geq q_1 \cdots q_k \geq 2^{k^2}$, which is impossible because $c < 1$. Now, suppose that

$$(4q_1) \cdots (4q_k) < x \leq (4q_1) \cdots (4q_{k+1}).$$

It implies that $2^{k^2+2k} < x \leq 2^{(k+2)^2+2(k+1)}$ and then $2^{k^2} = x^{1+o(1)}$. For the lower bound we have, $\mathcal{B}_{\overline{q},c}(x) \geq |\mathcal{P}_k| = \pi(2^{ck^2}/k^2) - \pi(2^{c(k-1)^2}/(k-1)^2) = 2^{ck^2(1+o(1))} = x^{c+o(1)}$. For the upper bound we have $\mathcal{B}_{\overline{q},c}(x) \leq \pi(2^{c(k+1)^2}/(k+1)^2) = 2^{ck^2(1+o(1))} = x^{c+o(1)}$. $\square$

2.1. **Proof of Theorems 1.1 and 1.2.** To prove Theorem 1.1 we simply observe that Proposition 1, iv) implies that $1 - c < \frac{c}{1-c}$, which does not hold for $c = \frac{3-\sqrt{5}}{2}$. Thus, $\mathcal{B}_{\overline{q},c}$ is a Sidon sequence for this value of c.

To prove Theorem 1.2 let us consider the sequence $\mathcal{B}^*_{\overline{q},c} = \{b_p\}_{p \in \mathcal{P}^*}$ where the numbers b_p are defined as above but where $\mathcal{P}^*$ are the survived primes after removing a thin subset to avoid the presence of some repeated sums. More precisely, we define $\mathcal{P}^* = \bigcup_{k_1} (\mathcal{P}_{k_1} \setminus \mathcal{R}_{k_1})$ where the removed set $\mathcal{R}_{k_1}$ consist of the primes in $\mathcal{P}_{k_1}$ dividing some integer of some of the sets S_{k_2,k_1}, $k_2 < \sqrt{\frac{c}{1-c}} k_1$, defined by

$$S_{k_2,k_1} = \left\{ s_1 Q_1 + s_2 Q_2 : \ 1 \leq |s_i| \leq \frac{2^{c(k_1^2+k_2^2)}}{k_1 k_2 Q_i}, \ i = 1, 2 \right\}$$

where $Q_1 = q_1 \cdots q_{k_2}$ and $Q_2 = q_{k_2+1} \cdots q_{k_1}$. We claim that for any $\overline{q}$ and $c = \sqrt{2} - 1$, the sequence $\mathcal{B}^*_{\overline{q},c} = \{b_p\}_{p \in \mathcal{P}^*}$ is an infinite Sidon sequence with $\mathcal{B}^*_{\overline{q},c}(x) = x^{\sqrt{2}-1+o(1)}$.

Suppose now that $b_{p_1} + b_{p_2} = b_{p'_1} + b_{p'_2}$ with $p_i, p'_i \in \mathcal{P}^*_{k_i}$, $i = 1, 2$ and $\{b_{p_1}, b_{p_2}\} \neq \{b_{p'_1}, b_{p'_2}\}$. First we observe that Proposition 1, iv) implies that $k_2 < \sqrt{\frac{c}{1-c}} k_1$. Next, note that thanks to parts ii) and ii) of Proposition 1 we can write

$$p_1(p_2 - p'_2) = s_1 Q_1 + s_2 Q_2$$

for the nonzero integers $s_1 = \frac{p_1 p_2 - p'_1 p'_2}{Q_1}$ and $s_2 = \frac{(p'_1 - p_1) p'_2}{Q_2}$ with $|s_i| \leq \frac{2^{c(k_1^2+k_2^2)}}{k_1 k_2 Q_i}$, $i = 1, 2$. It implies that $p_1(p_2 - p'_2) \in S_{k_2,k_1}$ for some $k_2 < \sqrt{\frac{c}{1-c}} k_1$, so $p_1 \in \mathcal{R}_{k_1}$ and then $p_1 \notin \mathcal{P}^*$.

To prove that $\mathcal{B}^*_{\overline{q},c}(x) = x^{c+o(1)}$ we only need to show that $|\mathcal{R}_{k_1}| = o(|\mathcal{P}_{k_1}|)$.

We observe first that $0 \notin S_{k_2,k_1}$. Otherwise we would have that $s_1 Q_1 = -s_2 Q_2$ for some s_1, s_2 and, since $(Q_1, Q_2) = 1$, it would imply that $Q_2 \mid s_1$ and then, $2^{k_1^2} \leq Q_2 Q_1 \leq |s_1| Q_1 \leq \frac{2^{c(k_1^2+k_2^2)}}{k_1 k_2} < 2^{\frac{c}{1-c} k_1^2}$ which is not possible because $c < 1/2$.

We claim that for each $n \in S_{k_2,k_1}$ and k_1 large enough, there exist at most one $p \in \mathcal{P}_{k_1}$ dividing n. Otherwise, if $p, p' \mid n$ we would have that

$$\frac{2^{2c(k_1-1)^2}}{(k_1-1)^2} < pp' \le |n| < \frac{2^{c(k_1^2+k_2^2)+1}}{k_1k_2} < \frac{2^{\frac{c}{1-c}k_1^2+1}}{k_1^2\sqrt{c/(1-c)}},$$

which dos not hold for k_1 large enough since $2c > \frac{c}{1-c}$ for $c < 1/2$. Therefore, using the bound $|\mathcal{P}_{k_1}| \gg \frac{2^{ck_1^2}}{k_1^3}$ and the identity $\frac{2c}{1-c} - 1 = c$ when $c = \sqrt{2} - 1$ we have, for k_1 large enough, the wanted estimate,

$$|\mathcal{R}_{k_1}| \le \sum_{k_2^2 < \frac{c}{1-c}k_1^2} |S_{k_2,k_1}| \le \sum_{k_2^2 < \frac{c}{1-c}k_1^2} \frac{2^{2c(k_1^2+k_2^2)-k_1^2+1}}{k_1^2k_2^2} \ll \frac{2^{\left(\frac{2c}{1-c}-1\right)k_1^2}}{k_1^4} \ll \frac{|\mathcal{P}_{k_1}|}{k_1}.$$

3. Infinite B_h sequences. Proof of Theorem 1.3

In the following we shall use the same notation with only minor changes. Fix $c = \sqrt{(h-1)^2+1} - (h-1)$ and let $\mathcal{P} = \cup_k \mathcal{P}_k$ where

$$\mathcal{P}_k = \left\{ p \text{ prime } : \ 2^{c(k-1)^2\left(1-1/\sqrt{\log(k-1)}\right)} < p \le 2^{ck^2\left(1-1/\sqrt{\log k}\right)} \right\}.$$

For $p \in \mathcal{P}_k$, we define the integer

$$b_p = \sum_{1\le j\le k} x_j(p)(h^2q_1)\cdots(h^2q_{j-1}),$$

where $x_j(p)$ is the solution of the congruence

$$g_j^{x_j(p)} \equiv p \pmod{q_j}, \qquad (h-1)q_j + 1 \le x_j(p) \le hq_j - 1,$$

and $x_j(p) = 0$ for $j > k$.

The sequence $\mathcal{B}_{\overline{q},c} = \{b_p\}$ will be a B_h sequence if for any $l,\ 2 \le l \le h$ there not exists a repeated sum

$$\begin{gathered} b_{p_1} + \cdots + b_{p_l} = b_{p'_1} + \cdots + b_{p'_l} \\ \{b_{p_1}, \ldots, b_{p_l}\} \cap \{b_{p'_1}, \ldots, b_{p'_l}\} = \emptyset. \end{gathered} \tag{3.1}$$

The plan of the proof is to remove from $\mathcal{B}_{\overline{q},c}$ the largest element appearing in each such repeated sum to obtain a true B_h sequence.

The following proposition is just a generalization of Proposition 1.

Proposition 3. *If* (3.1) *holds then*

i) $p_i, p'_i \in \mathcal{P}_{k_i},\ i = 1, \ldots, l$ *for some* $k_l \le \cdots \le k_1$.

ii)
$$\begin{array}{rcll} p_1\cdots p_l & \equiv & p_1'\cdots p_l' & \pmod{q_1\cdots q_{k_l}} \\ p_1\cdots p_{l-1} & \equiv & p_1'\cdots p_{l-1}' & \pmod{q_{k_l+1}\cdots q_{k_{l-1}}} \quad \textit{if } k_l<k_{l-1} \\ \cdots & & & \\ p_1 & \equiv & p_1' & \pmod{q_{k_2+1}\cdots q_{k_1}} \quad \textit{if } k_2<k_1. \end{array}$$

iii) $k_l^2<\frac{c}{1-c}\left(k_1^2+\cdots+k_{l-1}^2\right)$.

iv) $q_1\cdots q_{k_1} \mid \prod_{i=1}^l \left(p_1\cdots p_i - p_1'\cdots p_i'\right)$.

Proof. The proof is similar to the proof of Proposition 1: Here k_r is the first j such that $x(p_1)+\cdots+x(p_l)\ge r((h-1)q_j+1)$. Part iii) is consequence of the first congruence of part ii). Part iv) is also a consequence of part ii). □

3.1. **End of the proof.** We write $\mathcal{B}_{\overline{q},c}=\bigcup_k \mathcal{B}(\overline{q},k)$, where $\mathcal{B}(\overline{q},k)=\{b_p:\ p\in\mathcal{P}_k\}$. Note that for any sequence of primes $\overline{q}=(q_1,\dots,q_j,\dots)$ such that $2^{2j-1}<q_j\le 2^{2^j+1}$ we have that $|\mathcal{B}(\overline{q},k)|=|\mathcal{P}_k|=2^{ck^2(1+o(1))}$. Since $2^{2j-1}<q_j\le 2^{2j+1}$ for all j, we can proceed as in the previous section to deduce that $\mathcal{B}_{\overline{q},c}(x)=x^{c+o(1)}$.

The sequence $\mathcal{B}_{\overline{q},c}$ may not be a B_h sequence. Some repeated sums may eventually appear. But if we remove the largest b_p involved in each repetition, the survived elements of $\mathcal{B}_{\overline{q},c}$ will be a true B_h sequence. We define

$$\mathcal{R}(\overline{q},k)=\{b_p\in\mathcal{B}(\overline{q},k): b_p \text{ is the largest involved in some equation (3.1)}\}$$

It is then clear that the sequence

$$\mathcal{B}'_{\overline{q},c}=\bigcup_K\left(\mathcal{B}(\overline{q},k)\setminus\mathcal{R}(\overline{q},k)\right)$$

is a B_h sequence. If in addition we have that $|\mathcal{R}(\overline{q},k)|=o(|\mathcal{B}(\overline{q},k)|)$, it will be easy to check that

$$\mathcal{B}'_{\overline{q},c}(x)\sim\mathcal{B}_{\overline{q},c}(x)=x^{c+o(1)}.$$

Thus, the proof of Theorem 1.3 will be completed if we find some sequence $\overline{q}$ such that $|\mathcal{R}(\overline{q},k)|=o(|\mathcal{B}(\overline{q},k)|)$.

For $2\le l\le h$ we write

$$\mathrm{Bad}_l(\overline{q},k_l,\dots,k_1)=\{(p_1,\dots,p_l'):\ p_i,p_i'\in\mathcal{P}_{k_i},\ i=1,\dots,l \text{ satisfying (3.1)}\}.$$

Next let us observe that each bad element $b_p\in\mathcal{R}(\overline{q},k)$ comes from some $(p_1,\dots,p_l')\in\mathrm{Bad}_l(\overline{q},k_l,\dots,k_1)$, $2\le l\le h,\ k_l\le\cdots\le k_1=k$. Thus,

$$\begin{aligned} |\mathcal{R}(\overline{q},k)| &\le \sum_{l=2}^h\sum_{k_l\le\cdots k_1=k}|\mathrm{Bad}_l(\overline{q},k_l,\dots,k_1)| \\ &\le hk^{h-1}\max_{\substack{2\le l\le h\\ k_l\le\cdots\le k_1=k}}|\mathrm{Bad}_l(\overline{q},k_l,\dots,k_1)|. \end{aligned} \tag{3.2}$$

It happens that we are not able to give a good upper bound for $|\mathrm{Bad}_l(\overline{q}, k_l, \dots, k_1)|$ for a concrete sequence $\overline{q} = (q_1, \dots, q_j, \dots)$, but we can do it in average. If the reader is familiarized with Ruzsa's work, the sequences $\overline{q}$ will play the same role as the parameter α in Ruzsa's construction.

We consider the probability space of the sequences $\overline{q} = (q_1, \dots, q_j, \dots)$ where each q_j is chosen at random uniformly between all the primes in $(2^{2j-1}, 2^{2^j+1}]$. In particular, using that $\pi(2^{2k+1}) - \pi(2^{2k-1}) \gg 2^{2k}/k$ we have that

$$\begin{aligned} \mu(\overline{q} : \ q_1, \dots, q_{k_1} \in \overline{q}) &= \prod_{k=1}^{k_1} \frac{1}{\pi(2^{2k+1}) - \pi(2^{2k-1})} \\ &\leq 2^{-k_1^2 + O(k_1 \log k_1)}. \end{aligned}$$

Thus, for a given $(p_1, \dots, p_l')$, we use Proposition 3, iv) and the estimate $d(n) = n^{O(1/\log\log n)}$ for the divisor function to deduce that

$$\begin{aligned} \mu(\overline{q} : \ (p_1, \dots, p_l') \in \mathrm{Bad}_l(\overline{q}, k_l, \dots, k_1)) &\leq \sum_{\substack{q_1, \dots, q_{k_1} \\ q_1 \cdots q_{k_1} | \prod_{i=1}^l (p_1 \cdots p_i - p_1' \cdots p_i')}} \mu(\overline{q} : \ q_1, \dots, q_{k_1} \in \overline{q}) \\ &\leq d\left(\prod_{i=1}^l (p_1 \cdots p_i - p_1' \cdots p_i')\right) 2^{-k_1^2 + O(k_1 \log k_1)} \\ &\leq 2^{-k_1^2 + O(k_1^2/\log k_1)}. \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbb{E}((p_1, \dots, p_l') : \ p_i, p_i' \in \mathcal{P}_{k_i}, \ i = 1, \dots, l \text{ satisfying (3.1)}) \\ \leq\ & 2^{-k_1^2 + O(k_1^2/\log k_1)} \#((p_1, \dots, p_l') : \ p_i, p_i' \in P_{k_i}\} \\ \leq\ & 2^{-k_1^2 + O(k_1^2/\log k_1)} 2^{2c\left(k_1^2 + \cdots + k_l^2\right) - 2ck_1^2/\sqrt{\log k_1}} \\ \leq\ & 2^{-k_1^2 + \frac{2c}{1-c}\left(k_1^2 + \cdots + k_{l-1}^2\right) - (2c + o(1))k_1^2/\sqrt{\log k_1}} \\ \leq\ & 2^{\left(-1 + \frac{2c(l-1)}{1-c}\right)k_1^2 - (2c + o(1))k_1^2/\sqrt{\log k_1}}. \end{aligned}$$

Using (3.2) we have

$$\mathbb{E}(|\mathcal{R}(\overline{q}, k)|) \leq 2^{\left(-1 + \frac{2c(h-1)}{1-c}\right)k^2 - (2c + o(1))k^2/\sqrt{\log k}}.$$

Finally we use that $|\mathcal{B}(\overline{q}, k)| = |\mathcal{P}_k| \gg 2^{ck^2(1 - 1/\sqrt{\log k})}/k^2$ and that $-1 + \frac{2c(h-1)}{1-c} - c = 0$ for $c = \sqrt{(h-1)^2 + 1} - (h-1)$ to obtain

$$\begin{aligned} \mathbb{E}\left(\sum_k \frac{|\mathcal{R}(\overline{q}, k)|}{|\mathcal{B}(\overline{q}, k)|}\right) &\leq \sum_k k^2 2^{\left(-1 + \frac{2c(h-1)}{1-c} - c\right)k_1^2 - (c + o(1))k_1^2/\sqrt{\log k_1}} \\ &\leq \sum_k k^2 2^{-(c + o(1))k_1^2/\sqrt{\log k_1}}. \end{aligned}$$

Since the series is convergent we have that for almost all sequences $\overline{q}$ the series

$$\sum_k \frac{|\mathcal{R}(\overline{q},k)|}{|\mathcal{B}(\overline{q},k)|}$$

is convergent. Therefore, for any of these $\overline{q}$ we have that $|\mathcal{R}(\overline{q},k)| = o(|\mathcal{B}(\overline{q},k)|)$, which is what we wanted to prove.

4. An alternative construction

The theorems proved in this paper could have been proved using the following alternative construction, which, although has the same flavor than the one described in section 2, it uses irreducible polynomials in $\mathbb{F}_2[X]$ instead of the prime numbers.

Let $\overline{q} = \{q_j\}$ any infinite sequence of irreducibles polynomials in $\mathbb{F}_2[X]$ of degree $\deg(q_j) = 2j-1$. Fix $c,\ 0 < c < 1/2$ and for each $k \geq 2$, let

$$\mathcal{P}_k = \left\{p \text{ irreducible polynomials in } \mathbb{F}_2[X] :\ c(k-1)^2 < \deg p \leq ck^2\right\}.$$

Consider the sequence $\mathcal{B}_{\overline{q},c} = \{b_p\}$ where, for any $p(X) \in \mathcal{P}_k$ (we write $p := p(X)$ for short), the element b_p is defined by

$$b_p = \sum_{1\leq j\leq k} x_j(p) 2^{j^2+2j} \tag{4.1}$$

and $x_j(p)$ is the solution of the polynomial congruence

$$X^{x_j(p)} \equiv p(X) \pmod{q_j(X)}, \qquad 2^{2j-1}+1 \leq x_j(p) \leq 2^{2j}-1.$$

Let us define $x_j(p) = 0$ for $j > k$. Note that $b_p = \sum_{1\leq j\leq k} x_j(p)(4\cdot 2^1)\cdots(4\cdot 2^{2j-1})$ and then the digits of b_p in the base $4\cdot 2^1, \dots, 4\cdot 2^{2j-1}, \dots$ are $b_p = x_k x_{k-1}\cdots x_2 x_1$.

Using that the number of irreducible polynomials of degree j in $\mathbb{F}_2[X]$ is $\gg 2^j/j$, we can deduce easily that in this case we also have $\mathcal{B}_{\overline{q},c}(x) = x^{c+o(1)}$. Propositions 1 and 3 also work here, except that now the congruences are polynomial congruences.

It is then very easy to adapt the proofs of Theorems 1.1, 1.2 and 1.3 to this new construction, where perhaps, the less known ingredient may be the upper bound $d(r(X)) \leq 2^{O(n/\log n)}$ for the number of the divisors of a polynomial $r(X) \in \mathbb{F}_2[X]$ of degree n (see [4]).

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi, *A dense infinite Sidon sequence*, European J. Combin. 2 (1981), 1–11.

[2] J. Cilleruelo and R. Tesoro, *Infinite dense B_h sequences*, preprint. Arxiv.

[3] T. Gowers, *What are dense Sidon subsets of* $\{1, 2, \dots, n\}$ *like?*, Gowers's Webblog, July 2012

[4] Ph. Piret, *On the number of divisors of a polynomial over* $GF(2)$, Applied algebra, algorithmics and error-correcting codes (Toulouse, 1984), 161–168, Lecture Notes in Comput. Sci., 228, Springer, Berlin, 1986.

[5] I. Ruzsa, *An infinite Sidon sequence*, J. Number Theory 68 (1998).

Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM) and Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain
E-mail address: franciscojavier.cilleruelo@uam.es